

# Investigating the transmission risk of infectious disease outbreaks through the Aotearoa Co-incidence Network (ACN): a population-based study

S.M. Turnbull,<sup>a,b,\*</sup> M. Hobbs,<sup>c,d</sup> L. Gray,<sup>e</sup> E.P. Harvey,<sup>b,f</sup> W.M.L. Scarrold,<sup>b,g,h</sup> and D.R.J. O'Neale<sup>a,b</sup>

<sup>a</sup>Department of Physics, University of Auckland, Auckland, New Zealand

<sup>b</sup>Te Pūnaha Matatini, University of Auckland, Auckland, New Zealand

<sup>c</sup>GeoHealth Laboratory, University of Canterbury, Christchurch, Canterbury, New Zealand

<sup>d</sup>School of Health Sciences, University of Canterbury, Christchurch, Canterbury, New Zealand

<sup>e</sup>Department of Primary Health Care and General Practice, University of Otago, Wellington, New Zealand

<sup>f</sup>M.E. Research, Takapuna, Auckland, New Zealand

<sup>g</sup>Mathematical Institute, University of Oxford, Oxford, United Kingdom

<sup>h</sup>Institute for New Economic Thinking, Oxford Martin School, Oxford, United Kingdom

## Summary

**Background** The emergence and re-emergence of infectious diseases presents a significant challenge to public health and broader society. This study utilises novel nationwide data to calculate the transmission risk and potential inequity of infectious disease outbreaks through use of network analysis.

**Methods** Nationwide employment and education microdata (~4.7 million individuals in Aotearoa New Zealand) were used to develop the Aotearoa Co-incidence Network (ACN). The ACN considers connections generated when individuals are employed at the same workplaces or enrolled at the same schools. Through forms of network analysis, connections between geospatial areas can be established and provide proxy measures of infectious disease transmission risk. The ACN was also overlaid with nationwide population vulnerability data based on the number of older adults (>65 years) and individuals with long-term health conditions.

**Findings** We identify areas that have both high potential transmission risk (i.e., highly connected) and high vulnerability to infectious diseases. Community detection identified geographic boundaries that can be relevant to the application of regional restrictions for limiting infectious disease transmission.

**Interpretation** Integrating novel network science and geospatial analytics provides a simple way to study infectious disease transmission risk and population vulnerability to outbreaks. Our replicable method has utility for researchers globally with access to such data. It can help inform equitable preparation for, and responses to infectious disease outbreaks.

**Funding** This project was funded by the Health Research Council of New Zealand (20/1442) and from the NZ Government via Ministry for Business Innovation and Employment and Department of Prime Minister and Cabinet.

**Copyright** © 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Background

Infectious diseases are emerging and re-emerging at a rate not seen before.<sup>1</sup> Considered alongside declining vaccination coverage,<sup>2</sup> increasing antimicrobial resistance<sup>3</sup> and inequity<sup>4</sup> these present some of the greatest challenges of the 21<sup>st</sup> century. The current COVID-19 infectious disease pandemic has restricted usual

activities such as immunisation programmes and routine hospital care in many countries across the globe.<sup>5</sup> Consequently, a better understanding of potential infectious disease transmission will help future policy and research respond to new and re-emerging infectious diseases. With the increased collection of large complex data sets, such as those collected for administrative purposes, there is increased opportunity to understand potential transmission pathways prior to the development of an outbreak. In order to realise this opportunity, efforts are needed to provide generalisable methods that can make use of these complex forms of data. The current study seeks to provide one such

\*Corresponding author: Dr Steven Martin Turnbull, Post-Doctoral Research Fellow, University of Auckland, Auckland, New Zealand.

E-mail address: [s.turnbull@auckland.ac.nz](mailto:s.turnbull@auckland.ac.nz) (S.M. Turnbull).

The Lancet Regional Health - Western Pacific 2022;20: 100351

Published online 1 January 2022

<https://doi.org/10.1016/j.lanwpc.2021.100351>

## Research in Context

### *Evidence before this study*

Network analysis has a strong history in epidemiology, as outlined by Danon et al.<sup>7</sup> It has increasingly been used to quantify transmission risk for infectious diseases, based on analysis of contact, or interaction networks. Network analysis can be combined with empirical data sources to help infer transmission contexts and patterns of transmission between geospatial areas. Zhang et al.<sup>34</sup> and Irini et al.<sup>25</sup> used geo-located mobile phone data to infer mobility, while Liu et al. used existing case data from cases of COVID-19 in Hong Kong. Irini et al.<sup>25</sup> found that workplaces served as important contexts for infectious disease transmission in Ireland, while Munday et al.<sup>21</sup> conducted network analysis on census data to show how schools serve as important contexts for infectious disease transmission in the UK.

### *Added value of this study*

We utilised an internationally unique source of individual-level microdata, the Integrated Data Infrastructure, to create a publicly available tool named the Aotearoa Coincidence Network (ACN). The ACN is a unique network containing over 600,000 connections between geospatial areas in Aotearoa New Zealand (NZ), generated through the combination of census, workplace, and education data, for ~4.7 million individuals. The connections in the ACN represent the number of potential shared incidences between individuals in workplaces and schools. We use the ACN to estimate the transmission risk pathways due to interactions in the contexts of work and education. We employ a novel network analysis approach to identify important, highly connected geospatial areas across NZ. More specifically, we use network centrality as a proxy measure of transmission risk. This is based on both the *strength* and the *structure* of connections between different areas. We apply community detection to reveal spatially contiguous regional clusters within the ACN. Such regions can be used to guide the implementation of non-pharmaceutical interventions such as “lock-downs”. Building on the growing body of research using network analysis to study infectious diseases, the ACN is a rare example of how networks can be used to represent potential transmission pathways across a whole country. The ACN is a powerful tool as it can be used to inform public health responses to emerging pandemics, without the need for detailed case data or additional forms of data collection. The analytical methods presented in this study are easily replicable and will have value to researchers across international contexts. Our study also demonstrates how the ACN can be combined with existing nationwide data on vulnerability in terms of existing long-term health conditions to explore the equity of infectious disease outbreaks.

### *Implication of all the available evidence*

We identify several geospatial areas that carry a high risk of infectious disease transmission based on the

number of connections that individuals share through workplaces and schools. The ACN also revealed broader regional clusters and communities that tend to be more densely connected. Community detection outlined a distinct set of regional geographic boundaries that can inform the application of any regionally specific strategies to mitigate transmission spread such as mandated use of masks, workplace/school closures, or increased testing should any outbreak of infectious disease occur. By integrating the ACN with nationwide geospatial data on population vulnerability, the current study also identifies areas at high risk of transmission that are also highly vulnerable to an outbreak. Findings from the ACN were provided to the NZ government to assist in the response to the 2020 and 2021 outbreaks of COVID-19. The current study will be of interest to researchers, policymakers and communities located across the world, and in particular the Western Pacific, and between disciplines. All data, code, and an associated web application (<https://stur600.shinyapps.io/aotearoa-coincidence-network/>), is freely available and can be used as a tool to explore transmission risk and vulnerability at a fine-grained level.

method of understanding potential transmission risk, by employing a novel form of network analysis to explore how regions are connected through shared interaction contexts.

Network analysis is a useful tool for understanding the transmission of infectious diseases<sup>6</sup> with a strong history in the field of epidemiology.<sup>7</sup> Network structures consist of collections of nodes and links, where nodes can represent individuals or entities, and links represent the relationship between them. Networks can be complex: they may contain two or more node types (multipartite), or multiple types of links (multigraphs). As an example, nodes representing individuals may connect to nodes representing workplaces. In such a network, workers would share indirect connections to their colleagues through a common workplace node. Network analysis has been much used in the study of infectious diseases, whether that be sexually transmitted diseases<sup>8</sup> livestock diseases<sup>9</sup> and more recently, COVID-19.<sup>10</sup> Transmission naturally takes on the structure of an interaction network where connections shared between individuals may be used to represent links in a potential chain of transmission.<sup>7</sup> Such networks can be used to predict the distribution of infections following an outbreak.

Despite the value that can be gained through studying infectious disease transmission through network analysis, doing so can be challenging. Social network analysis is often employed to investigate infectious diseases spread through contacts;<sup>11</sup> however, this often requires considerable resources and time-consuming surveys<sup>12</sup> or expensive equipment such as wireless sensor technology.<sup>13</sup> Emerging sources of large and complex data, often collected for administrative purposes,

have potential to help mitigate the challenges of these forms of data collection by inferring interaction contexts.<sup>14</sup> Such data can be used to address pressing global health challenges such as infectious disease outbreaks without the need for additional data collection.<sup>15</sup> Moreover, the rapid digitalisation of health, education and broader social systems and significant improvements in data processing and storage capabilities have allowed researchers to access several sources of novel data not previously available, such as linked nationwide data on households, education, and employment.<sup>16</sup>

Efforts continue to be made to model infectious disease transmission using complex systems approaches.<sup>14</sup> With advances in computing power, the utility of networks to analyse numerous sources of data have increased dramatically,<sup>17,18</sup> and studies employing network analysis to understand disease transmission patterns are increasingly popular.<sup>10,19,20</sup> Liu and colleagues<sup>10</sup> demonstrated how networks can help uncover spatiotemporal transmission patterns, using existing data collected on COVID-19 cases across different districts in Hong Kong. Importantly, they argue that understanding connectivity across geospatial areas can help guide public health responses to future pandemics. Munday and colleagues<sup>21</sup> used census data from the UK to create a network of schools connected to households, and through this approach were able to assess the potential impact of school closures on COVID-19 transmission. Fewer studies have combined transmission risk data with an understanding of population vulnerability within particular geographical areas. This adds important contextual information about which areas may be most affected and informs an equitable public health response.<sup>22</sup> For instance, evidence has suggested certain individuals, such as those with medical conditions may have a higher risk of severe illness from COVID-19.<sup>22</sup>

Our study contributes a novel methodological approach combining different sources of nationwide data to develop the Aotearoa Co-incidence Network (ACN). The ACN is a network of geographic areas that may be used as a proxy measure of potential infectious disease transmission. The ACN considers the connections shared between different geospatial areas through shared workplaces and schools. Both contexts serve as “hubs” which connect many individuals and have been identified as key contexts where disease transmission takes place.<sup>21,23–25</sup> Previous studies, such as that conducted by Munday and colleagues<sup>21</sup> have used networks where schools are connected to households to study how schools may impact on infectious disease transmission. While these forms of analysis are valuable, the current study will provide an alternative method that can contribute insights into potential transmission pathways without the need to consider the location of schools or workplace. In addition to providing the

means to study potential transmission pathways in a way that preserves anonymity of schools and workplaces, a simplified network structure focusing on the geographic areas in which people inhabit can provide additional important insights in terms of potential infectious disease transmission. The following section will outline how the ACN is constructed, how it can be used to derive proxy measures of geospatial transmission risk, and how it may be combined with other sources of data, such as health vulnerability, to assess the equity of potential outbreaks.

## Method

### Study design

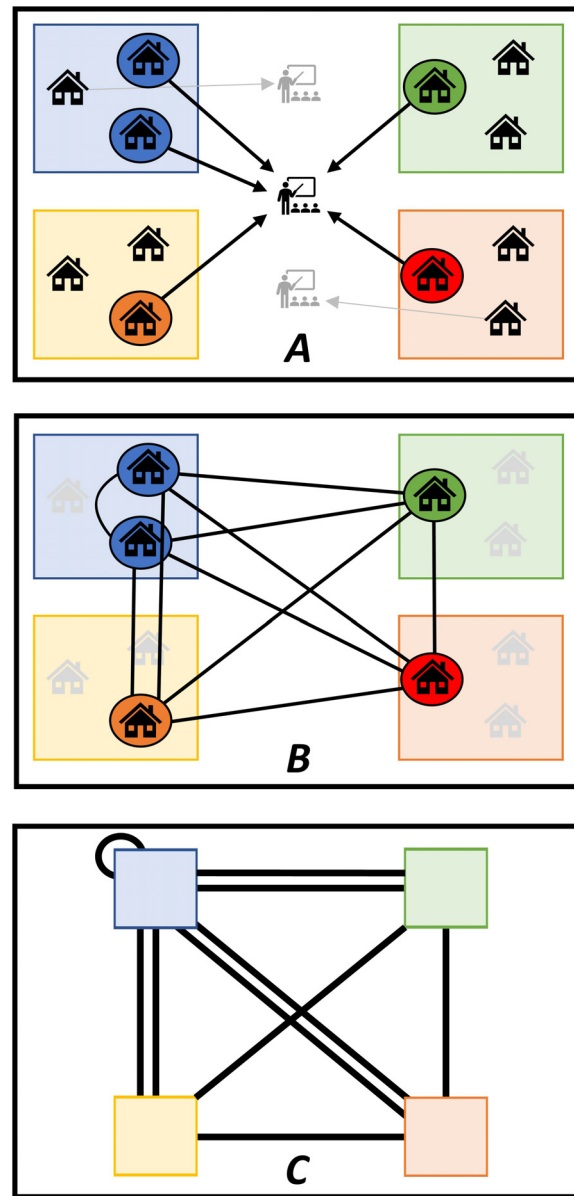
This was a nationwide, cross-sectional and geospatial study in Aotearoa New Zealand (NZ) approved by Statistics NZ (reference: MAA2020-36). All code and data used in the current are publicly available and can be accessed at: <https://gitlab.com/tpm-public-projects/aotearoa-connection-network>. Network analysis was carried out in R<sup>26</sup> using the *igraph* package.<sup>27</sup>

### Data source: Integrated Data Infrastructure (IDI)

Nationwide data were obtained from the Integrated Data Infrastructure (IDI) for ~4.7 million individuals present in NZ census 2018 records, along with their household (dwelling of usual residence in census 2018), school enrolments (current enrolment data from Ministry of Education) and place of employment (wages and salaries data from Inland Revenue). The IDI is a unique collection of individual-level linked microdata for people in NZ.<sup>28</sup> All data are linked and completely de-identified before being made available to researchers within a secure data lab.<sup>28</sup> The IDI environment is only accessible for researchers and projects approved by Statistics NZ. All extracted data are checked by Statistics NZ before release to ensure non-identifiability of individuals or entities covered by the aggregated data, this includes suppression of low counts and random rounding of satisfactorily high counts.

### Developing the Aotearoa Co-incidence Network (ACN)

The Aotearoa Co-incidence Network (ACN) is constructed by generating a network consisting of two sets of nodes (a bipartite network), where dwellings are linked to those schools/workplaces where inhabitants of the dwelling are enrolled/employed (Figure 1: Panel 1A). We then “project” onto the dwelling nodes in the network to obtain a network consisting of only dwellings, with links connecting dwellings when inhabitants share a workplace or school (Figure 1: Panel 1B). In addition to making the network simpler, the projection focuses the analysis on areas



**Figure 1.** A simplified depiction of the development of the Aotearoa Co-incidence Network (ACN) for the case of five people from five different dwellings attending the same school. Panel 1A represents how dwellings in one area can share connections with dwellings in different areas through a shared interaction context (e.g., school). Panel 1B shows how the network can be simplified by focusing on the indirect connections shared by dwellings through that interaction context. This is simplified further by aggregating across geospatial areas (Panel 1C).

corresponding to individuals' place of usual residence, making it possible to directly compare with data on regional vulnerabilities. The projected network is then aggregated into connections between the geospatial area units that contain the dwellings, defined by Statistics NZ as Statistical Area 2 (SA2), to produce the ACN (Figure 1: Panel 1C). Importantly, the connections shared between SA2s in the ACN can be considered potential pathways by which infectious diseases may be transmitted.

SA2s are part of the Statistical Standard for Geographic Areas 2018 (SSGA2018)<sup>29</sup> and can contain around up to 4,000 residents (rural SA2s may contain fewer than 1000 residents).<sup>29</sup> We exclude SA2s located outside of NZ Territorial Authorities (TAs), leaving 2,147 SA2s in the final network, connected by 669,878 links. Each link has an associated value corresponding to the total number of combinations of inhabitants from the pair of SA2s who are co-employed or co-enrolled at a workplace or school.

### Identifying connected communities

We analyse the patterns of connections in the ACN using community detection<sup>30</sup> which serves to partition the network into different communities of SA2s. Communities represent clusters of SA2s that tend to be more strongly connected to areas within the same community than they are to SA2s in other communities. Partitioning the ACN into these communities is a useful technique for coarse graining the network to a manageable size<sup>17</sup> and providing an overview of how different areas are connected. We use a method of community detection that employs modularity maximisation to identify the most robust communities. In simple terms, modularity maximisation identifies communities that result in the highest ratio of the number of links within groups, relative to those between groups. This method can be applied to the ACN based on all types of regional connections, or for different types of workplace and school connections separately. This is valuable, since the importance of different types of connection can vary depending on context (e.g., schools may be closed, only essential workplaces open). We tested various modularity maximisation community detection methods, which resulted in similar communities and modularity scores.

### Transmission risk and vulnerability

We calculate a proxy measure of transmission risk in the ACN using PageRank centrality. Centrality measures provide a valuable metric in identifying “important” nodes in a network based on the number and patterns of links shared between different nodes. There are a number of different metrics for computing different types of centrality<sup>17</sup> each containing different underlying assumptions about what “importance” means. The simplest centrality measure, degree centrality, simply counts the number of connections (the degree) to each node; betweenness centrality counts the number of paths on a network, between arbitrary pairs of nodes, that pass through any specific nodes. In this study, we use PageRank - a flavour of eigenvector centrality.<sup>17,31,32</sup>

PageRank centrality is an appropriate centrality measure for infection on networks as it captures the concept of node “importance” in the sense of the probability of a random spreading process on a network visiting a specific node. PageRank centrality has wide usage in studies of infectious disease transmission, including studies investigating the importance of a geographic location in human flow networks<sup>33</sup> as well as research seeking to identify facilities most at risk from Bovine Viral Diarrhea Virus.<sup>34</sup> For the current study, PageRank centrality is implemented using the *igraph* package<sup>27</sup> in R<sup>26</sup> by treating the links between SA2s as bidirectional (i.e., we assume transmission could originate from either of a pair of SA2s). PageRank then provides a proxy for

transmission risk by considering both the structure and strength of connections between SA2s<sup>17</sup> not only in the connections from one SA2 to its neighbours, but also the connections from those neighbours to next-nearest neighbour SA2s, and so on. An SA2 will be deemed high risk if the summed risk of its neighbours is high. This covers both the case when a SA2 has many neighbours and when a SA2 has a few high-risk neighbours.

We overlaid the spatial estimates of transmission risk, derived from the ACN, with data detailing the vulnerability of each geospatial area at SA2 level. Vulnerability data is taken from Wiki and colleagues<sup>35</sup> and is based on the proportion of the population with long-term health conditions (LTCs). LTCs were sourced from the National Minimum Dataset for the period 2011–2016. LTCs included cancer, cardiovascular conditions, diabetes, renal conditions, and respiratory illnesses. Due to the strong association between COVID-19 vulnerability and age, each of these factors is combined with the percentage of individuals over 65 years to generate a composite score of vulnerability for each SA2.

To explore and visualise SA2s with high potential transmission risk and population vulnerability, we created a bivariate legend (Figure 2) which cross-referenced these sets of data, with each dimension split into tertiles (Figure 2). For instance, category ‘A’ would be low on both potential transmission risk and health vulnerability while ‘F’ would be high potential transmission risk, but low health vulnerability and ‘D’ would be high in terms of health vulnerability but low potential transmission risk. Perhaps the most important category is ‘I’ which indicates areas that rank in the highest tertile for both potential transmission risk and health vulnerability. To aid exploration into the vulnerability we also report distributions at the level of TA areas — a much larger geospatial unit, which represent the second tier of local government below Regional Council areas.<sup>36</sup>

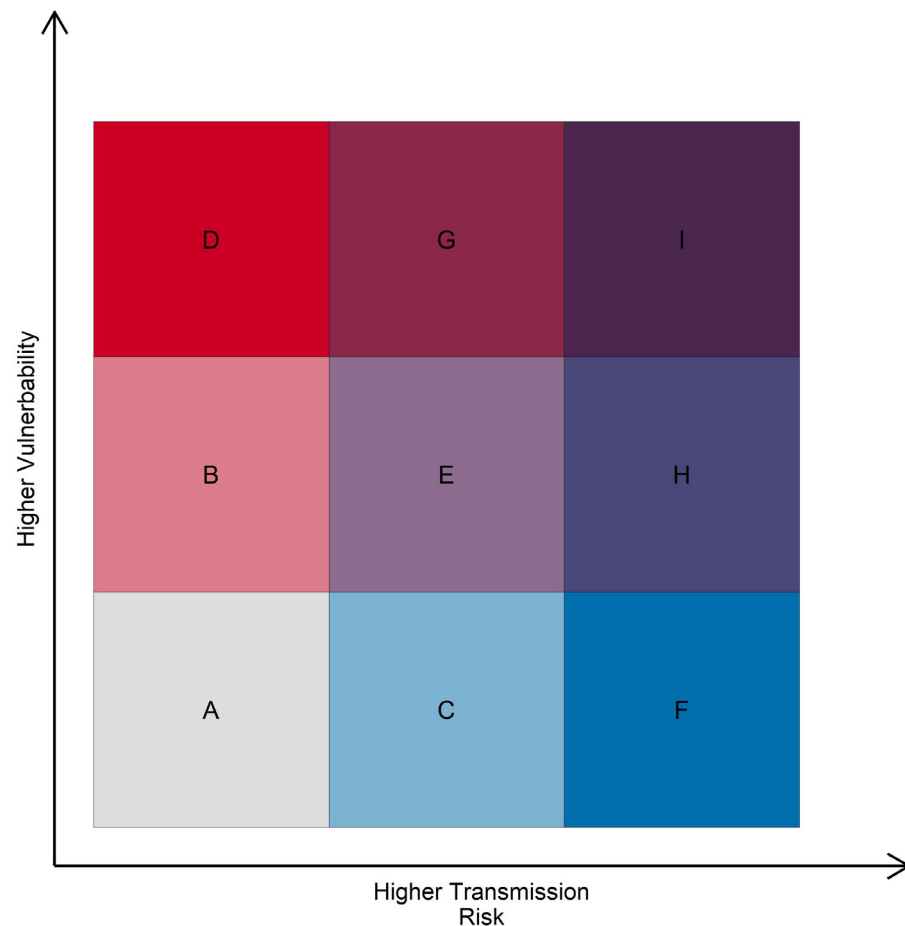
### Role of Funding Source

This project was funded by the Health Research Council of New Zealand (20/1442) and from the New Zealand Government via Ministry for Business Innovation and Employment and Department of Prime Minister and Cabinet contracts for modelling advice on responding to COVID-19 in Aotearoa New Zealand. The funding source had no role in study design, data collection, data analysis, interpretation, or writing of this report

## Results

### Aotearoa Co-incidence Network

Figure 3 presents the connections shared between central SA2s across NZ cities from the ACN. We observe that most connections are made to SA2s near the



**Figure 2.** Bivariate legend of SA2s based on potential transmission risk and health vulnerability. Note: grey (category A) represents areas with low potential transmission risk and low vulnerability; blue (category F) represents areas with high potential transmission risk and low vulnerability; red (category D) represents areas with low potential transmission risk and high vulnerability; Purple (category I) represents areas with high potential transmission risk and high vulnerability.

selected SA2, with fewer connections as distance increases. However, the ACN also reveals more detailed patterns. For example, Wellington Central has a significant number of connections to the north of the South Island. Detailed findings are available in a public-facing, interactive web application (see <https://stur600.shinyapps.io/aotearoa-coincidence-network/>).

#### Connected communities

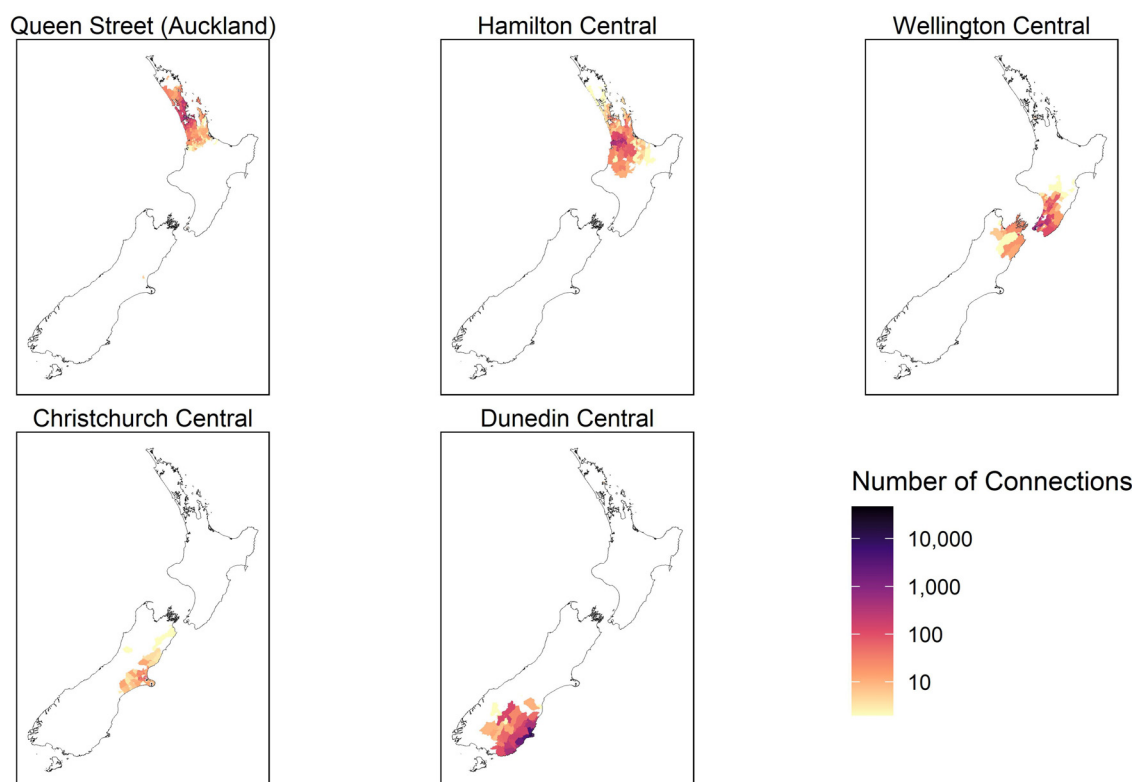
Through use of community detection, we also identified the geographic communities that are present in the ACN. We present the findings for all types (both work and education) of connections in the ACN. Geographic communities are indicated by the different colours in Figure 4A. The communities identified strongly reflect established local government (TA) boundaries with some exceptions. For example, the Auckland community extends further south than the Auckland TA boundary (see Figure 4B). Communities based solely on

school and workplace connections are presented in supplementary material (Figure S1).

#### Transmission risk

Using the connections in the ACN we calculate the PageRank centrality of each SA2 to identify areas that can be considered most at risk in terms of potential infectious disease transmission. We tested various measures of centrality (supplementary material, Figure S2), but present PageRank since it captures both strength of connections and the effect of relative risk from neighbouring connected areas. Figure 5A shows the distribution of PageRank scores across NZ at the level of SA2. Areas with increased potential transmission risk (i.e., higher PageRank centrality) tend to be found in clusters. This is reflected in the observation that PageRank establishes a clear partition between urban and rural areas (supplementary material, Figure S3) such that urban areas are more likely to have higher PageRank scores and thus





**Figure 3.** The number of connections from central SA2s in Auckland, Hamilton, Wellington, Christchurch, and Dunedin. Colour gradient represents the number of connections from the specified SA2 to other SA2s.

higher potential transmission risk. We find that distributions vary across different regions (Figure 5B), with some appearing normal (e.g., Auckland), and others appearing multimodal (e.g., Hamilton). This highlights the diversity in regional transmission risk and the importance of understanding transmission risk at finer-grained geospatial levels.

#### Combining transmission risk and population vulnerability

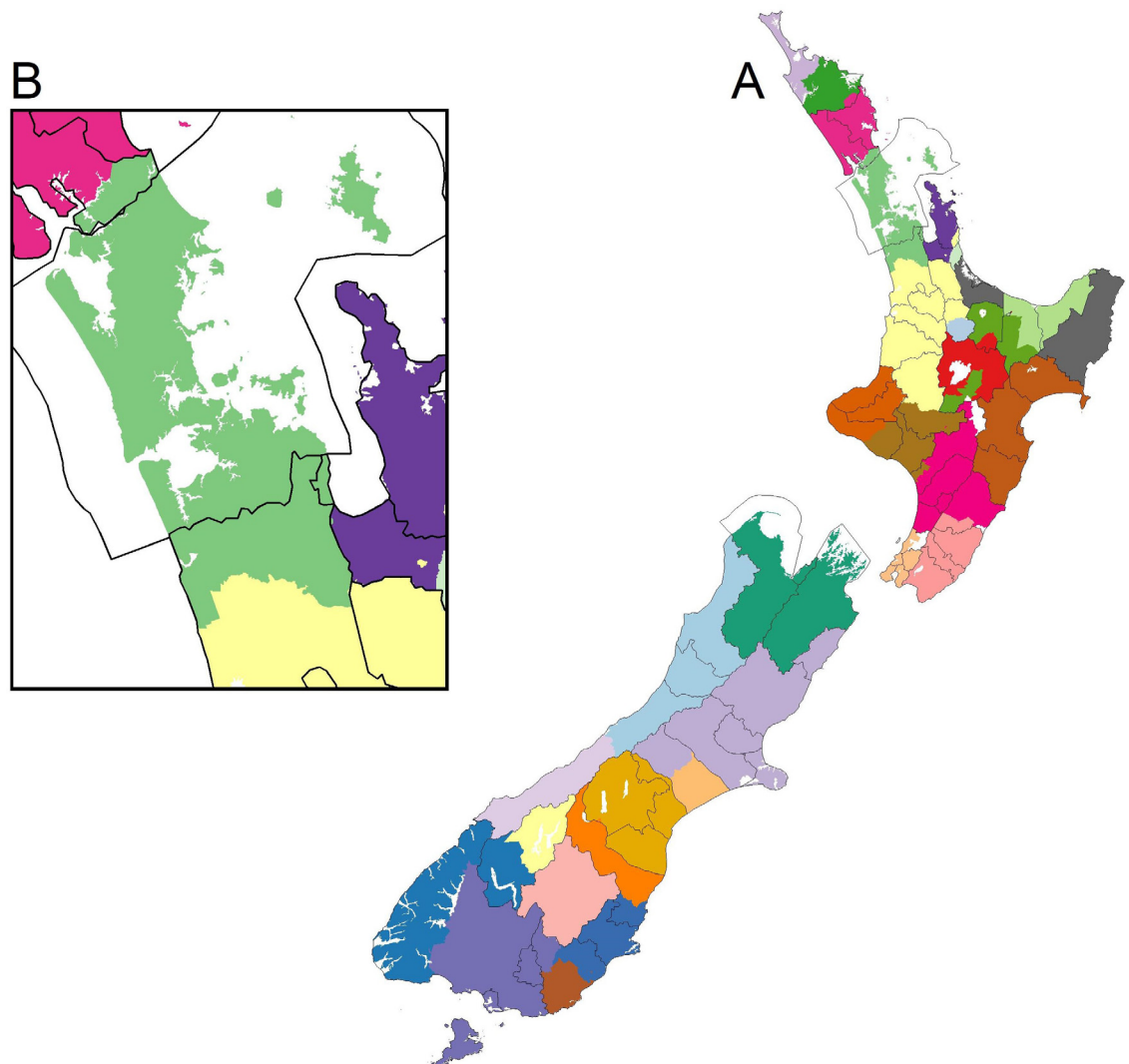
Figure 6A combines potential transmission risk (PageRank centrality) with estimates of health vulnerability<sup>35</sup> for each SA2 across NZ. Significant areas of the North and South Island were categorised as low risk tertiles (grey shaded areas, category A) for both potential transmission risk and health vulnerability. Despite this, several areas of the North and South Island were classified as high health vulnerability but low potential transmission risk (red shaded areas, category D). Large parts of Wellington and Hamilton were classified as high potential transmission risk but low health vulnerability (blue shaded areas, category F). High potential transmission risk and high health vulnerability were identified in several areas of the North and South Island including in Christchurch, Hamilton, Dunedin and Auckland. Dunedin tends to have a higher proportion of SA2s that are

potentially high risk and high vulnerability, while Wellington is mainly composed of SA2s that have high potential transmission risk but low vulnerability (Figure 6B).

#### Discussion

Understanding the connections between people and places is important for modelling potential infectious disease transmission.<sup>10,19</sup> The current study used a unique and replicable methodology based in network science to estimate potential transmission risk for communicable disease across geospatial regions. We combined several sources of nationwide data in NZ on dwellings, schools, and workplaces to produce the ACN. The ACN is highly informative for understanding potential infectious disease transmission pathways to calculate proxy measures of infectious disease transmission risk.

Our study presented a novel method for determining geographical boundaries, generated through community detection on the ACN. The utility of the ACN therefore extends to informing application of non-pharmaceutical interventions, helping to curb the spread of infectious disease<sup>37</sup> widely employed in the Western Pacific and around the world.<sup>38–40</sup> This includes approaches such as such as defining regional

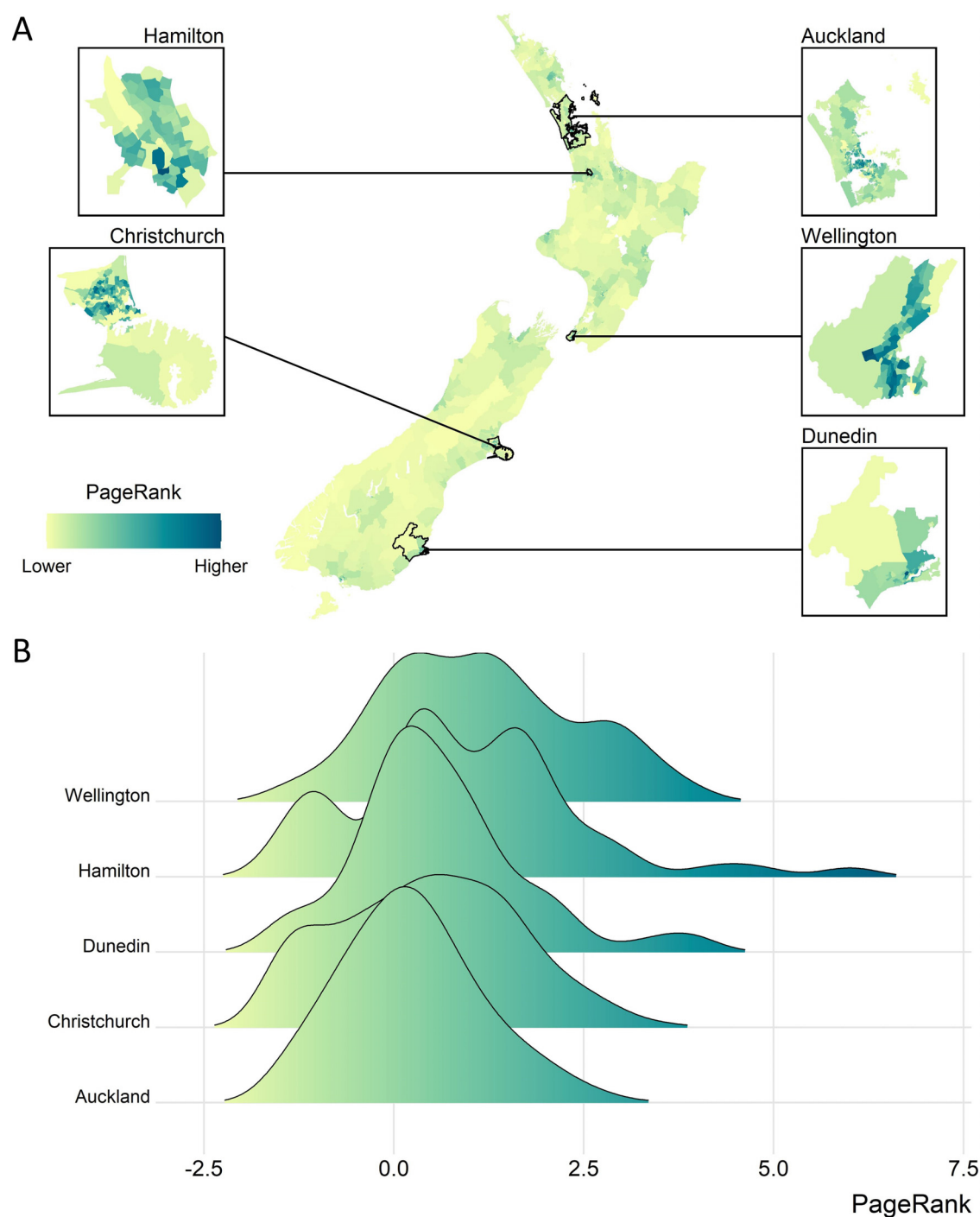


**Figure 4.** Communities detected in the ACN for the whole of New Zealand (Panel 4A), and for the Auckland region specifically (Panel 4B). Territorial Authority boundaries are outlined in black.

boundaries or mandating use of masks, workplace/school closures or increased testing.<sup>41</sup> Such findings will inform the allocation of resources across NZ and identify communities that are at high risk of transmission and help NZ government make decisions to contain the spread of the community outbreaks of COVID-19.<sup>42</sup> Prior to the development of the ACN, during the Auckland regional lockdown in February 2020, police checkpoints were extended further south.<sup>43</sup> (Figure S3B) to allow for the number of commuters that needed to cross the Auckland TA boundary, even with reduced numbers of people working on-site during the lockdown period. The ‘Auckland community’ detected in the ACN (Figure S4A) provides a better fit to this adapted boundary compared to the original TA boundary, which may validate the ACN’s capability and impact as a planning tool.

Our study shows that centrality, or the extent of connectedness between areas, is higher in urban areas. We confirm findings observed in other studies that a higher density of inhabitants and potential close contact between people, for instance in urban areas<sup>44</sup> may provide ideal conditions for the rapid spread of infectious diseases.<sup>45</sup> Despite this, other evidence shows that risk of transmission often varies across continents, across a country, and even within a city.<sup>10,45</sup> While sufficiently aggregated to preserve privacy, the ACN reveals at a fine geographical scale where variations in potential transmission risk may exist and shows that areas with high levels of centrality tend to be clustered together geospatially. This observation reflects Tobler’s first rule of geography:<sup>46</sup> “everything is related to everything else, but near things are more related than distant things.”

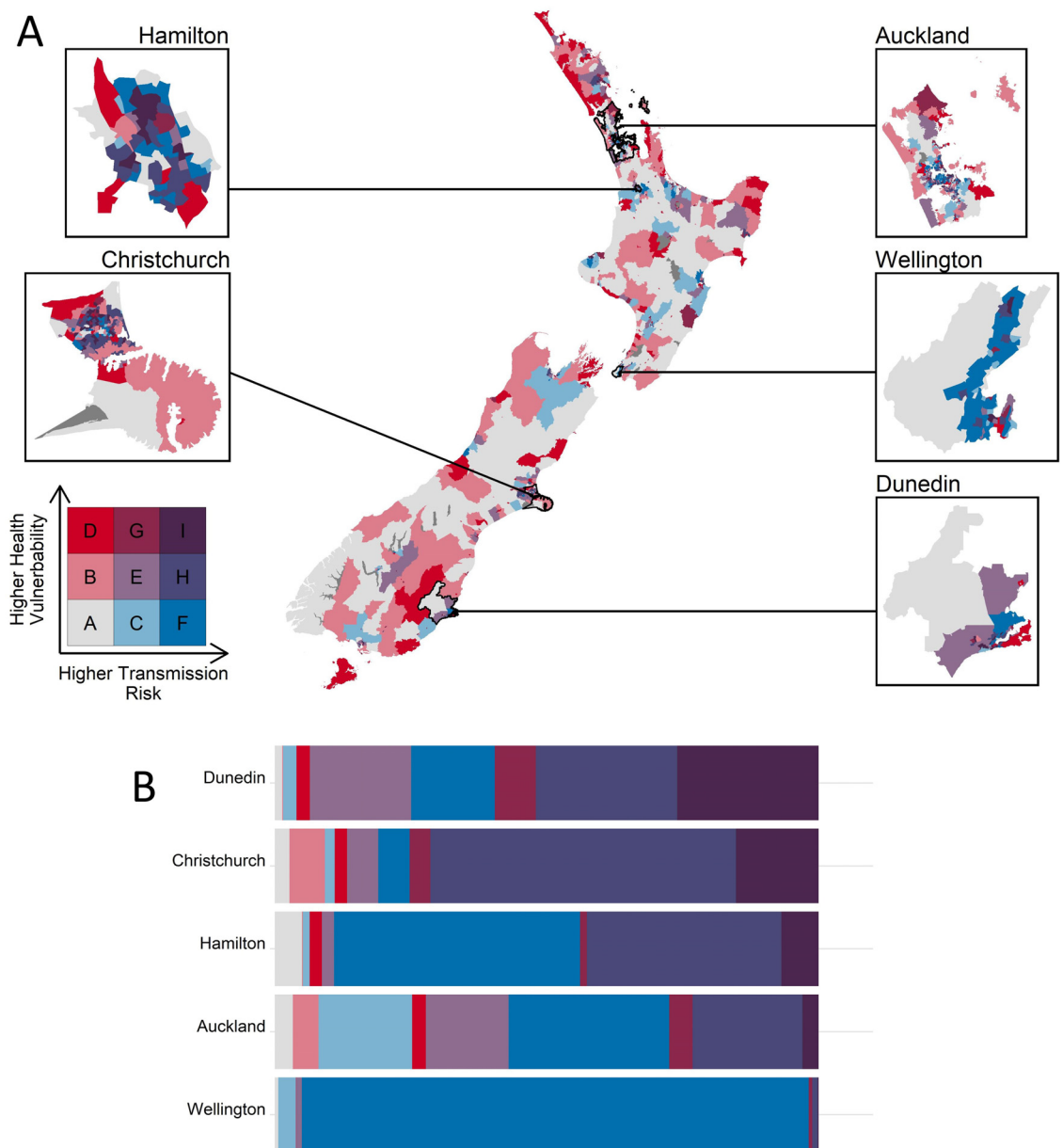




**Figure 5.** PageRank centrality scores across Aotearoa New Zealand (Panel 5A) and associated distributions across SA2s located in the major Territorial Authorities of Auckland, Christchurch, Hamilton, Dunedin, and Wellington (Panel 5B). The colour gradient across both Panel 5A and 5B represent PageRank centrality, where higher PageRank provides an indication of higher potential transmission risk.

In addition to identifying areas with high potential transmission risk, it is also important to account for vulnerable populations residing within such areas.<sup>47</sup> Some

areas may be more vulnerable to infectious disease, in terms of age, health conditions and access or lack of access to healthcare services.<sup>48</sup> Our study highlights



**Figure 6.** Combined potential transmission risk and population vulnerability based on long-term health conditions (Panel 6A) and distribution of risk for each SA2 located in the major territorial authorities of Dunedin, Christchurch, Hamilton, Auckland and Wellington (Panel 6B). Colours indicate levels of potential transmission risk and health vulnerability, as represented in [Figure 2](#).

several geographic areas at risk of infectious disease transmission containing a high proportion of the population with long-term health conditions. Previous research highlights which areas may be vulnerable<sup>35</sup> where super-spreading events may have occurred<sup>44</sup> or areas that may have inequitable access to vaccination centres<sup>48</sup> but seldom have such data been combined with a proxy for transmission risk as provided by the ACN.

There are several policy implications from the ACN. Considering that infectious diseases are emerging and

re-emerging at a rate not seen before<sup>1</sup> our methodology, based on novel network and geospatial data science, provides a simple yet effective way to estimate potential transmission risk to any infectious disease outbreak. Importantly, it is replicable and has application for any researcher with access to sources of linked data. An enhanced understanding of how different areas may be connected and the vulnerability of the population within those areas are important for governments and communities in mitigating the spread of infectious diseases.<sup>10</sup>

The utility of the ACN was further evidenced in its use as a resource for informing the NZ government's response to COVID-19, with technical reports provided to the government during outbreaks of COVID-19 in 2021.<sup>49</sup>

A strength of the ACN is its whole population approach, using relatively comprehensive nationwide data from the IDI (n~4.7 million), but limitations within the IDI data do exist.<sup>28</sup> The New Zealand census population is limited in its representation of the current day NZ population, with it being smaller than 2021 estimates.<sup>50</sup> However, the current study can be easily reproduced with future census data, and by other researchers around the world with similar data. Another limitation is that there is a slight temporal mismatch in the census data used to build the ACN (2018) and the data on vulnerability (2011-2016), while the proportion of individuals with long-term health conditions within the vulnerability data is also just one aspect of vulnerability. Future work could extend the ACN to look at more detailed information regarding the types of workplaces and schools where individuals share an interaction context and consider more information regarding population vulnerabilities. For example, healthcare and hospitality workplaces may have more risky interactions compared to the information technology sector<sup>51</sup> while secondary schools may serve as larger hubs in a network, compared to primary schools.<sup>21</sup>

## Conclusion

In a time where infectious diseases are continuing to emerge, efforts to make increased use of existing sources of data are especially valuable. We outline a novel, reproducible methodology where we represent shared interaction contexts through workplaces and schools as the ACN. Our results at a broad level show that the areas of most risk within the ACN are urban areas, and areas of high risk are often clustered within close proximity to one another. However, the ACN provides the granularity to explore risk and vulnerability at a detailed level that can help inform regional responses to outbreaks of infectious disease. The ACN is a powerful tool for informing responses to outbreaks of infectious disease, and its reproducible methodology makes it relevant for researchers across the world.

## Disclaimer

Access to the data used in this study was provided by Stats NZ under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the author, not Stats NZ or individual data suppliers. The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994 for statistical purposes. Any discussion

of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

## Contributors

SMT, MH, EH, DRJO conceived this study and developed the methodology. SMT carried out the formal analysis, and SMT and WMLS curated the data. SMT, EH, and WMLS contributed to the software and validation of results. SMT and MH contributed to the visualisations. SMT, MH, LG, EH, and DRJO contributed to the writing, review and editing of the manuscript. SMT, MH, LG, EH, and DRJO contributed to funding acquisition. DRJO led the project administration. All authors gave final approval of the version to be published.

## Data Sharing Statement

All code and data used in the current are publicly available and can be accessed at: <https://gitlab.com/tpm-public-projects/aotearoa-connection-network>. A public-facing Shiny web application can also be accessed at: <https://stur600.shinyapps.io/aotearoa-coincidence-network/>

## Declaration of Competing Interest

This project was funded by the Health Research Council of New Zealand (20/1442) and from the New Zealand Government via Ministry for Business Innovation and Employment and Department of Prime Minister and Cabinet contracts for modelling advice on responding to COVID-19 in Aotearoa New Zealand. The authors have no conflict of interests to declare.

## Funding

This project was funded by the Health Research Council of New Zealand (20/1442) and from the New Zealand Government via Ministry for Business Innovation and Employment and Department of Prime Minister and Cabinet contracts for modelling advice on responding to COVID-19 in Aotearoa New Zealand.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.lanwpc.2021.100351](https://doi.org/10.1016/j.lanwpc.2021.100351).

## References

- 1 Fauci AS. Emerging and reemerging infectious diseases: the perpetual challenge. *Academic Medicine* 2005;80:1079–85.
- 2 Marek L, Hobbs M, McCarthy J, et al. Investigating spatial variation and change (2006–2017) in childhood immunisation coverage in New Zealand. *European Journal of Public Health* 2020;264:113292.

- 3 Laxminarayan R, Boeckel T van, Frost I, et al. The Lancet Infectious Diseases Commission on antimicrobial resistance: 6 years later. *The Lancet Infectious Diseases* 2020;20:e51–60.
- 4 Beaglehole R. The challenge of health inequalities. *The Lancet* 2006;367:559–60.
- 5 Venkatesan P. Re-emergence of infectious diseases associated with the past. *The Lancet Microbe* 2021;2:e140.
- 6 Wey T, Blumstein DT, Shen W, Jordán F. Social network analysis of animal behaviour: a promising tool for the study of sociality. *ANIMAL BEHAVIOUR* 2008. <https://doi.org/10.1016/j.anbehav.2007.06.020>.
- 7 Danon L, House TA, Read JM, Keeling MJ. Social encounter networks: collective properties and disease transmission. *Journal of The Royal Society Interface* 2012;9:2826–33.
- 8 Wylie JL, Jolly A. Patterns of Chlamydia and Gonorrhea Infection in Sexual Networks in Manitoba, Canada. *Sexually transmitted diseases* 2001;28:14–24.
- 9 Chaters GL, Johnson PCD, Cleaveland S, et al. Analysing livestock network data for infectious disease control: an argument for routine data collection in emerging economies. *Philosophical Transactions of the Royal Society B* 2019;374:20180264.
- 10 Liu Y, Gu Z, Liu J. Uncovering transmission patterns of COVID-19 outbreaks: A region-wide comprehensive retrospective study in Hong Kong. *EClinicalMedicine* 2021;36:100929.
- 11 Cauchemez S, Bhattarai A, Marchbanks TL, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proceedings of the National Academy of Sciences* 2011;108:2825–30.
- 12 Mossong J, Hens N, Jit M, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Disease. *PLoS Medicine* 2008. <https://doi.org/10.1371/journal.pmed.0050074>.
- 13 Salathé M, Kazandjieva M, Lee JW, Lewis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. *PNAS* 2010; 105. <https://doi.org/10.1073/pnas.1009094108>.
- 14 Mistry D, Litvinova M, y Piontti AP, et al. Inferring high-resolution human mixing patterns for disease modeling. *Nature Communications* 2021;12. <https://doi.org/10.1038/s41467-020-20544-y>.
- 15 Fuller D, Buote R, Stanley K. A glossary for big data in population and public health: discussion and commentary on terminology and research methods. *Journal of Epidemiology and Community Health* 2017. <https://doi.org/10.1136/jech-2017-209608>. published online Sept 16.
- 16 Oldroyd RA, Hobbs M, Campbell M, et al. Progress Towards Using Linked Population-Based Data For Geohealth Research: Comparisons Of Aotearoa New Zealand And The United Kingdom. *Applied Spatial Analysis and Policy* 2021;14. <https://doi.org/10.1007/s12061-021-09381-8>.
- 17 Newman M. Networks. Oxford university press, 2018.
- 18 Vespignani A. Twenty years of network science. *Nature* 2018;558:528–9.
- 19 Lee HA, Alves LGA, Nunes Amaral LA. Spreader events and the limitations of projected networks for capturing dynamics on multi-partite networks. *Physical Review E* 2021;103. <https://doi.org/10.1103/PhysRevE.103.022320>.
- 20 Martín-Calvo D, Aleta A, Pentland A, Egido YMEM. Effectiveness of social distancing strategies for protecting a community from a pandemic with a data-driven contact network based on census and real-world mobility data. 2020; published online March. [https://connection.mit.edu/sites/default/files/publication-pdfs/Preliminary\\_Report\\_Effectiveness\\_of\\_social\\_distance\\_strategies\\_COVID-19](https://connection.mit.edu/sites/default/files/publication-pdfs/Preliminary_Report_Effectiveness_of_social_distance_strategies_COVID-19).
- 21 Munday JD, Sherratt Katharine, Meakin S, et al. Implications of the school-household network structure on SARS-CoV-2 transmission under school reopening strategies in England. *Nature Communications* 2021;12. <https://doi.org/10.1038/s41467-021-22213-0>.
- 22 Ministry of Health NZ. COVID-19: Advice for higher risk people. 2021.
- 23 Lee J, Kim M. Estimating the number of Korean workers at high risk of COVID-19 infection. *Epidemiology and Health* 2020; e2020051.
- 24 Chernozhukov V, Kasahara H, Schrimpf P. The association of opening K12 schools with the spread of COVID-19 in the United States: County-level panel data analysis. *Proceedings of the National Academy of Sciences* 2021; 118. <https://doi.org/10.1073/pnas.2103420118>.
- 25 Irini F, Kia AN, Shannon D, Jannusch T, Murphy F, Sheehan B. Associations between mobility patterns and COVID-19 deaths during the pandemic: A network structure and rank propagation modelling approach. *Array* 2021;11:100075.
- 26 R Core Team. R: A language and environment for statistical computing. 2013.
- 27 Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, complex systems* 2019.
- 28 Milne BJ, Atkinson J, Blakely T, et al. Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI). *International Journal of Epidemiology* 2019;48. <https://doi.org/10.1093/ije/dyz014>.
- 29 Stats NZ. Statistical Area 2 2018 (generalised). 2018; published online Sept.
- 30 Edler D, Rosvall M. The MapEquation software package. 2013.
- 31 Bonacich P. Power and Centrality: A Family of Measures. *American Journal of Sociology* 1987;92. <https://doi.org/10.1086/228631>.
- 32 Perra N, Fortunato S. Spectral centrality measures in complex networks. *Physical Review E* 2008;78. <https://doi.org/10.1103/PhysRevE.78.036107>.
- 33 Zhang Y, Wang L, Zhu JJH, Wang X. The spatial dissemination of COVID-19 and associated socio-economic consequences. 2021.
- 34 Hirose S, Notsu K, Ito S, Sakoda Y, Isoda N. Transmission Dynamics of Bovine Viral Diarrhea Virus in Hokkaido, Japan by Phylogenetic and Epidemiological Network Approaches. *Pathogens* 2021;10. DOI: [doi.org/10.3390/pathogens10080922](https://doi.org/10.3390/pathogens10080922).
- 35 Wiki J, Marek L, Hobbs M, Kingham S, Campbell M. Understanding vulnerability to COVID-19 in New Zealand: a nationwide cross-sectional study. *Journal of the Royal Society of New Zealand* 2021;51. <https://doi.org/10.1080/03036758.2021.1900294>.
- 36 Stats NZ. Statistical standard for geographic areas 2018. 2017.
- 37 Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 2020;584:257–61.
- 38 Gathergood J, Guttman-Kenney B. The English Patient: Evaluating Local Lockdowns Using Real-Time COVID-19 & Consumption Data. *Covid Economics* 2021. <https://doi.org/10.2139/ssrn.3798666>.
- 39 Ren X. Pandemic and lockdown: a territorial approach to COVID-19 in China, Italy and the United States. *Eurasian Geography and Economics* 2020;61:423–34.
- 40 Caselli Mauro; Fracasso ASS. From the lockdown to the new normal: An analysis of the limitations to individual mobility in Italy following the Covid-19 crisis. 2021.
- 41 Chu DK, Akl EA, Duda S, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* 2020;395:1973–87.
- 42 Unite Against COVID-19. History of the COVID-19 Alert System. 2021.
- 43 New Zealand Police. Update on Auckland checkpoints (with maps). 2021.
- 44 Huang J, Kwan M-P, Kan Z. The superspreading places of COVID-19 and the associated built-environment and socio-demographic features: A study using a spatial network framework and individual-level activity data. *Health & Place* 2021;72:102694.
- 45 Neiderud C-J. How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology & Epidemiology* 2015;5:27060.
- 46 Tobler WR. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 1970;46. <https://doi.org/10.2307/143141>.
- 47 Baker MG, Wilson N, Blakely T. Elimination could be the optimal response strategy for covid-19 and other emerging pandemic diseases. *bmj* 2020; 371.
- 48 Whitehead J, Carr PA, Scott N, Lawrenson R. Spatial inequity in distribution of COVID-19 vaccination services in Aotearoa. *medRxiv* 2021. 2021.08.26.21262647.
- 49 Harvey EP, Hobbs M, Looker J, O'Neale DRJ, Scarrold W, Turnbull SM. Exploring COVID-19 Transmission Risk and Vulnerability Through the Aotearoa Co-incidenceNetwork (ACN). 2021.
- 50 Stats NZ. Estimated population of NZ. 2021.
- 51 Lan F-Y, Wei C-F, Hsu Y-T, Christiani DC, Kales SN. Work-related COVID-19 transmission in six Asian countries/areas: A follow-up study. *PLoS One* 2020;15:e0233588.